REGION-ORIENTED COMPRESSION OF MULTISPECTRAL IMAGES BY SHAPE-ADAPTIVE WAVELET TRANSFORM AND SPIHT

M. Cagnazzo, G. Poggi, L. Verdoliva

Dip. di Ing. Elettronica e Telecomunicazioni
Università Federico II di Napoli
via Claudio, 21 – 80125 Napoli, Italy
{cagnazzo,poggi,verdoliv}@unina.it

A. Zinicola

Lab. Naz. Comunicazioni Multimediali
CNIT
via Diocleziano, 328 – 80125 Napoli, Italy
andrea.zinicola@cnit.it

ABSTRACT

We present a new technique for the compression of remote-sensing hyperspectral images based on wavelet transform and zerotree coding of coefficients. In order to improve encoding efficiency, the image is first segmented in a small number of regions with homogeneous texture. Then, a shape-adaptive wavelet transform is carried out on each region, and the resulting coefficients are finally encoded by a shape-adaptive version of SPIHT. Thanks to the segmentation map (sent as a side information) region boundaries are faithfully preserved, and selective encoding strategies can be easily implemented. In addition, by-now homogeneous region textures can be more efficiently encoded.

Key-words: Multispectral-image compression, wavelet transform, SPIHT, object-oriented segmentation.

1. INTRODUCTION

Remote-sensing multispectral and hyperspectral images portray large areas of the Earth at high spatial resolution and, increasingly, on a large number of spectral bands. As a consequence, they require huge resources for transmission to the ground station, archival, and diffusion to the end users. For this reason, much attention has been devoted in recent years to the compression of such images, as testified by the many papers appeared in the literature. The major focus has been on transform coding, based on discrete cosine transform (DCT), Karhunen-Loeve transform (KLT), or discrete wavelet transform (DWT), followed by a suitable quantization step.

Some of the most popular schemes are based on hybrid transforms that treat differently spectral and spatial redundancies. In [1], for example, a KLT is carried out in the spectral domain, followed by a two-dimensional (2d) DCT in the spatial domain. A similar scheme is considered in [2] with the DWT playing the role of the DCT. The rationale behind the hybrid transform is the different nature of the dependencies in the spatial and spectral domain. In fact, along the spectral dimension, the signal depends almost exclusively on the pixel land cover and hence, if only a few land covers are present in the image, the KLT works fairly well. In the spatial dimensions, on the contrary, the signal depends on the scene geometry and is less predictable, with many discontinuities near region boundaries. In this case, the KLT becomes far from optimal, and DCT and DWT do a better job.

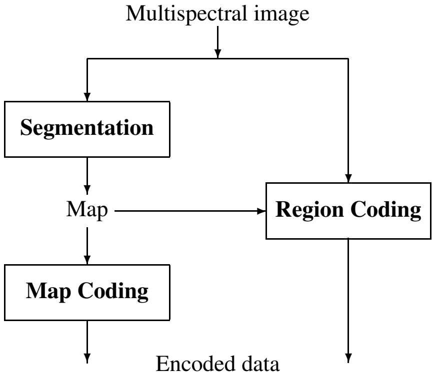

Fig. 1. Region-based coding scheme.

These considerations shed light on the fact that the transform coding approach, by itself, is ill suited to compress this kind of images, and a deeper modeling effort is necessary to fully exploit the data redundancies. In particular, it should be recognized that a remote-sensing image is often composed by a small number of regions with homogeneous parameters (land-cover, texture, etc.). If such regions are recognized and extracted, they can be separately encoded, according to the general scheme of Fig.1, with a likely gain in coding efficiency. The cost of side information, on the other hand, is typically negligible, because a single segmentation map must be encoded for all the bands of the image.

Of course, to implement region-based coding, several challenging problems must be addressed, concerning the segmentation, the lossless coding of the map, and the lossy

0-7803-8554-3/04/$20.00 ©2004 IEEE.

coding of the region textures. In our research, the first results of which are reported in this paper, we resort to the wavelet transform because of its simplicity and good compaction performance, followed by the SPIHT algorithm [3], an efficient and flexible zerotree-based coding technique. Shape-adaptive versions of both the WT and SPIHT have been implemented, in order to encode regions of arbitrary shape. In next section we describe in some detail the various tools proposed in the coding scheme, that is, the segmentation algorithm, the lossless map coding technique, the shape-adaptive wavelet transform, shape-adaptive SPIHT, and the rate-allocation strategy. Then, in Section 3, we present and comment some example experimental results.

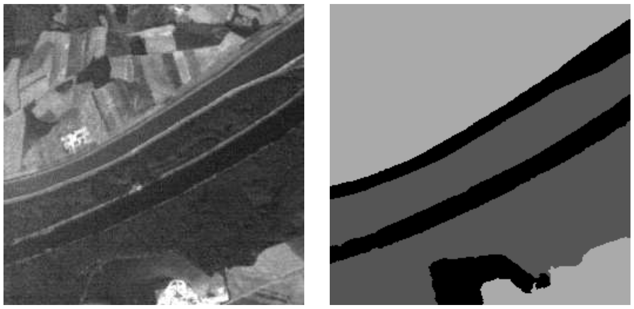
(a) (b)

Fig. 2. Original image (band 7) and segmentation map

2. THE CODING TOOLS

2.1. Segmentation

A meaningful segmentation of the image is of central importance for the success of a region-based coding scheme, but it is also a complex and largely unsolved problem. In our application we have two requirements: on one hand, we want each region to be formed by pixels of the same class, so as to exhibit homogeneous statistics and increase the efficiency of subsequent encoding. On the other hand, we would like to segment the image in a small number of large regions, in order to have a simple map to encode, and to use shape-adaptive transforms on nice regular shapes.

Such requirements are partially contrasting, since remote-sensing images typically present a large number of small regions and isolated points because of the sensor noise. A relatively smooth map can be obtained by resorting to the Bayesian approach, so as to include prior information in the segmentation process In particular, we used the technique proposed in [4], based on a tree-structured Markov random field image model, which proved to be accurate and fast. In addition, a few morphological operations were carried out in order to eliminate small regions and smooth out contours. An example of the results is given in Fig.2, where we show a band of the original 256×256-pixel image (a), and the segmentation map obtained with the MRF-based technique with morphological filtering (b). Although the final map comprises only seven regions, not all of them are really homogeneous, as appears for example in the upper-left region, which is composed of different kinds of fields and even some woods. This will likely affect the final performance, and the best compromise between smoothness and homogeneity is still an open question, to investigate in future research.

2.2. Map coding

After the segmentation map has been processed to smooth contours and erase small regions, it can be coded without loss of information at a limited cost. In the overall research project we are developing a contour tracer and encoder, in order to represent efficiently the map and extract at the same time significant geometric information for subsequent higher-level applications, such as data-mining. For the time being, however, we use a simpler scheme based on adaptive arithmetic coding. The K-ary map, with K the number of regions, is first converted into a sequence of $K-1$ binary maps in order to reduce complexity. For each pixel, a causal 4-pixel neighborhood is used, with $3^4 = 81$ possible contexts (a third symbol is needed to signal out-of-map neighbors), reduced for simplicity to just five contexts based on the count of zeros and ones, that is *all-zeros*, *majority-of-zeros*, *all-ones*, *majority-of-ones* and *mixed*. For our smooth maps, the vast majority of pixels have an *all-zeros/all-ones* contexts, and are significantly compressed, leading to a negligible cost for map coding. Of course, this scenario could change in case a more faithful map is used.

2.3. Shape-adaptive wavelet transform

Once the original image has been segmented in a set of homogeneous regions, we compress the regions' textures by means of transform coding. In particular, we resort to the wavelet transform to exploit its many appealing properties, first of all its ability to work on the whole image and compact energy in a few coefficients, treating equally well smooth regions and discontinuities. Of course, the basic transform must be adapted to operate on regions with arbitrary shapes. A simple approach, already used in the well-known shape-adaptive DCT, is to flush the pixels to the left edge of the bounding box, apply 1d transform along the rows, then flush again the coefficients to the upper edge and transform the coefficient along the columns. This technique is very simple but changes the mutual position of pixels, thus reducing the correlation between transformed pixels and, in the end, the compaction ability of the transform. Another possibility is to extend, with various strategies, the signal in the bounding box outside the region of interest, but the obvious drawback is the increase, possibly large, in the number of coefficients to be later encoded.

Recently, however, a new shape-adaptive wavelet transform algorithm has been proposed by Li and Li (we refer to [5] for a detailed description) which overcomes all these problems. This transform operates "in place", exploiting the spatial locality of DWT coefficients. As a consequence, it preserves the spatial correlation, locality properties of wavelet transforms, and the self-similarity across subbands, which is at the core of zerotree-based coding algorithms. In addition, the number of coefficients after transform is the same as the number of pixels in the object, and no redundancy is introduced. Finally, for a rectangular region the transform reduces to the conventional wavelet transforms. For all these reasons Li and Li's algorithm was included in the MPEG-4 standard, and we use it in our coding scheme, with the global subsampling modality [5].

2.4. Shape-adaptive SPIHT

SPIHT (set partitioning in hierarchical trees) is a well-known zerotree-based algorithm for the progressive quantization of wavelet transform coefficients [3]. It is simple, intrinsically scalable, and very efficient, which is why it is one of the most popular techniques for the compression of images and, more recently, video. In addition, it can be readily modified to encode images of arbitrary geometry after a shape-adaptive WT.

We introduced only two major changes with respect to the basic algorithm. First, only "active" nodes, that is nodes belonging to the support of the SA-DWT of the object, should be considered while scanning a spatial orientation tree. This information is available at the decoder since the segmentation map is sent without loss. The second modification concerns the baseband, where coefficients cannot be grouped anymore in 2×2 squares, as in the original algorithm, since they are not always well defined anymore, and a single root is considered instead.

Even though the encoding algorithm is readily modified to account for the arbitrary geometry, its performance might well suffer because of elongated or fragmented shapes, since many bits could be spent to encode sparsely populated branches. This will be another key topic in future investigation.

2.5. Rate allocation

This is one of the most interesting and promising aspects of region-based coding, since the allocation of resources to the different regions could well be driven by the application or even by the user itself in order to privilege certain areas over the others. For example, in multitemporal images, one can choose to encode (or decode) at higher rates only the areas where temporal changes have been detected.

When the goal is only the minimization of the distortion, we resort to the optimal Lagrangian approach, progressively allocating resources to the various objects until the same slope is reached on all rate-distortion curves.

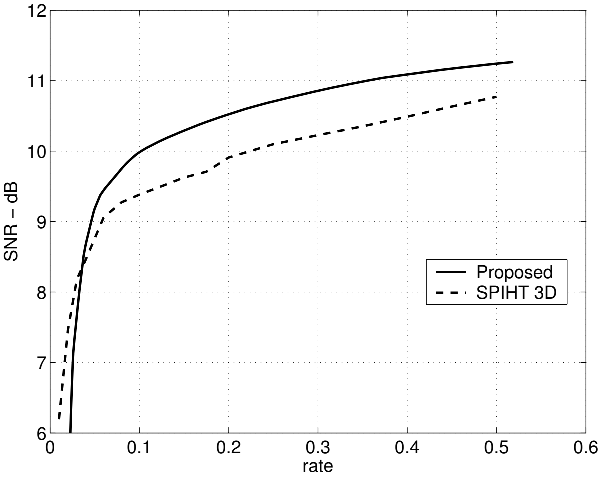

Fig. 3. Rate-distortion curves for proposed region-based techniques and reference (3d-SPIHT)

3. EXPERIMENTAL RESULTS

In this Section we present a few sample numerical results. Experiments are carried out on a 256x256 section of a 63-band hyperspectral image (GER sensor) one band of which was shown in Fig.2(a). The corresponding segmentation map of Fig.2(b) has been encoded as described in Section 2.2 with a cost of 0.080 bit/pixel. As for the wavelet transform, to obtain a good encoding efficiency, 5 levels of decomposition are used in the spatial domain (Daubechies 9/7 filters), but only 3 in the spectral domain (Haar filters), because only 8 bands of the image are jointly encoded here. Since a dyadic transform is not possible in this situation, we work first in the spectral domain (all 3 levels) and then in the spatial domain, obtaining a packet decomposition. Of course, the SPIHT encoder will follow this subband structure. After encoding all objects, the optimal rate allocation procedure was carried out to obtain the rate-distortion curve reported in Fig.3. For comparison purposes, we also report the rate-distortion curves of a state-of-the-art non shape-adaptive (or "flat" in the following) technique, 3d-SPIHT which is simply the flat counterpart of the proposed technique, with the same parameters.

At very low bit-rates shape-adaptive coding suffers because of the 0.01 bit/pixel/band (bpb) offset due to the side information for map coding. However, starting from 0.1 bpb it shows a significant improvement (always less than 1 dB, anyway) that persists also at higher rates although beyond 0.4 bpb it seems to start decreasing. It is worth noting that 0.5 bpb can be considered a relatively high rate for hyperspectral images, because of the high degree of correlation exhibited by the closely packed spectral bands.

The improvement in rate-distortion performance, however, is only one of the interesting features of the region-

Fig. 4. Band 7 at 0.1 bpb with flat coding

Fig. 5. Band 7 at 0.1 bpb with region-based coding

based approach, and maybe not the most relevant. In fact, segmentation singles out and encodes without loss of information the region boundaries, which largely contribute to the perceived image quality and, more important, are precious features for analysis and interpretation. With flat techniques, instead, these non-stationary components require a significant part of the overall bit budget and nonetheless are poorly represented at low bit-rates. To gain insight about this point, Fig.4 and Fig.5 show the test image encoded at 0.1 bpb by 3d-SPIHT, and by the proposed technique with optimal rate allocation, respectively. Here, region-based coding guarantees only a modest 0.6 dB increase in the SNR but the improvement in the overall perceived quality is dramatic: as expected, the edges are more accurately reproduced without the blurring effect observed with 3d-SPIHT; moreover, since many steep edges are removed thanks to segmentation, many resources can be saved and devoted to encode the textures in the regions' interior which in fact appear much more faithful.

Another interesting feature of the region-based approach, not shown here, is the freedom to allocate more resources to regions of interest. Of course, the other parts of the image will be barely encoded, but this could not be a problem for applications where the user is only interested on a particular region, maybe selected only by looking at the map.

In conclusion, the region-based approach seems very well suited to the compression of multispectral and hyperspectral images. Although there is certainly room for further improvements, the proposed coding scheme is already superior to state-of-the-art techniques, and represents a more flexible and feature-rich coding tool.

4. ACKNOWLEDGMENTS

The authors are grateful to Ing. Sofia Massascusa and Ing. Antonella Persichini for their help with the development of the software

5. REFERENCES

[1] J.A.Saghri, A.G.Tescher, J.T.Reagan, “Practical transform coding of multispectral imagery,” *IEEE Signal Processing Magazine*, pp.32–43, Jan 1995.

[2] P.L.Dragotti, G.Poggi, A.R.P.Ragozini, “Compression of multispectral images by three-dimensional SPIHT algorithm,” *IEEE Transactions on Geoscience and Remote Sensing*, pp.416–428, Jan. 2000.

[3] A.Said, W.A.Pearlman, “A new fast and efficient image codec based on set partitioning in hierarchical trees,” *IEEE Transactions on Circuits and Systems for Video Technology*, pp.243–250, June 1996.

[4] C.D'Elia, G.Poggi, G.Scarpa: “A tree-structured Markov random field model for bayesian image segmentation,” *IEEE Transactions on Image Processing*, pp.1259–1273, Oct. 2003.

[5] S.Li, W.Li, “Shape-adaptive discrete wavelet transforms for arbitrarily shaped visual object coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, pp.725–743, Aug.2000.